

ENTERED*See page 6*

1600

RECEIVED

MAY 15 2003

TECH CENTER 1600/2900

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/701,395A

DATE: 05/08/2003

TIME: 14:48:25

Input Set : A:\10817222.app

Output Set: N:\CRF4\05082003\I701395A.raw

```

3 <110> APPLICANT: CUNNINGHAM JR., FRANCIS X.
4   SUN, ZAIREN
6 <120> TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
7   METHODS OF USE THEREOF
9 <130> FILE REFERENCE: 108172-00022
11 <140> CURRENT APPLICATION NUMBER: 09/701,395A
12 <141> CURRENT FILING DATE: 2001-09-25
14 <150> PRIOR APPLICATION NUMBER: 09/088,724
15 <151> PRIOR FILING DATE: 1998-06-02
17 <150> PRIOR APPLICATION NUMBER: 09/088,725
18 <151> PRIOR FILING DATE: 1998-06-02
20 <160> NUMBER OF SEQ ID NOS: 62
22 <170> SOFTWARE: PatentIn Ver. 2.1
24 <210> SEQ ID NO: 1
25 <211> LENGTH: 1860
26 <212> TYPE: DNA
27 <213> ORGANISM: Arabidopsis thaliana
29 <220> FEATURE:
30 <221> NAME/KEY: CDS
31 <222> LOCATION: (109)..(1680)
33 <400> SEQUENCE: 1
34 acaaaaaggaa ataattagat tctcttttct gcttgcata ccttgataga acaatataac 60
36 aatggtgtaa gtcttctcgc tgtattcgaa attatttga ggaggaaa atg gag tgt 117
37                                     Met Glu Cys
38                                     1
40 gtt ggg gct agg aat ttc gca gca atg gcg gtt tca aca ttt ccg tca 165
41 Val Gly Ala Arg Asn Phe Ala Ala Met Ala Val Ser Thr Phe Pro Ser
42      5                      10                      15
44 tgg agt tgt cga agg aaa ttt cca gtg gtt aag aga tac agc tat agg 213
45 Trp Ser Cys Arg Arg Lys Phe Pro Val Val Lys Arg Tyr Ser Tyr Arg
46 20                      25                      30                      35
48 aat att cgt ttc ggt ttg tgt agt gtc aga gct agc ggc ggc gga agt 261
49 Asn Ile Arg Phe Gly Leu Cys Ser Val Arg Ala Ser Gly Gly Gly Ser
50      40                      45                      50
52 tcc ggt agt gag agt tgt gta gcg gtg aga gaa gat ttc gct gac gaa 309
53 Ser Gly Ser Glu Ser Cys Val Ala Val Arg Glu Asp Phe Ala Asp Glu
54      55                      60                      65
56 gaa gat ttt gtg aaa gct ggt ggt tct gag att cta ttt gtt caa atg 357
57 Glu Asp Phe Val Lys Ala Gly Gly Ser Glu Ile Leu Phe Val Gln Met
58      70                      75                      80
60 cag cag aac aaa gat atg gat gaa cag tct aag ctt gtt gat aag ttg 405
61 Gln Gln Asn Lys Asp Met Asp Glu Gln Ser Lys Leu Val Asp Lys Leu
62      85                      90                      95

```

RAW SEQUENCE LISTING

DATE: 05/08/2003

PATENT APPLICATION: US/09/701,395A

TIME: 14:48:25

Input Set : A:\10817222.app

Output Set: N:\CRF4\05082003\I701395A.raw

```

64 cct cct ata tca att ggt gat ggt gct ttg gat cat gtg gtt att ggt 453
65 Pro Pro Ile Ser Ile Gly Asp Gly Ala Leu Asp His Val Val Ile Gly
66 100 105 110 115
68 tgt ggt cct gct ggt tta gcc ttg gct gca gaa tca gct aag ctt gga 501
69 Cys Gly Pro Ala Gly Leu Ala Leu Ala Ala Glu Ser Ala Lys Leu Gly
70 120 125 130
72 tta aaa gtt gga ctc att ggt cca gat ctt cct ttt act aac aat tac 549
73 Leu Lys Val Gly Leu Ile Gly Pro Asp Leu Pro Phe Thr Asn Asn Tyr
74 135 140 145
76 ggt gtt tgg gaa gat gaa ttc aat gat ctt ggg ctg caa aaa tgt att 597
77 Gly Val Trp Glu Asp Glu Phe Asn Asp Leu Gly Leu Gln Lys Cys Ile
78 150 155 160
80 gag cat gtt tgg aga gag act att gtg tat ctg gat gat gac aag cct 645
81 Glu His Val Trp Arg Glu Thr Ile Val Tyr Leu Asp Asp Asp Lys Pro
82 165 170 175
84 att acc att ggc cgt gct tat gga aga gtt agt cga cgt ttg ctc cat 693
85 Ile Thr Ile Gly Arg Ala Tyr Gly Arg Val Ser Arg Arg Leu Leu His
86 180 185 190 195
88 gag gag ctt ttg agg agg tgt gtc gag tca ggt gtc tcg tac ctt agc 741
89 Glu Glu Leu Leu Arg Arg Cys Val Glu Ser Gly Val Ser Tyr Leu Ser
90 200 205 210
92 tcg aaa gtt gac agc ata aca gaa gct tct gat ggc ctt aga ctt gtt 789
93 Ser Lys Val Asp Ser Ile Thr Glu Ala Ser Asp Gly Leu Arg Leu Val
94 215 220 225
96 gct tgt gac gac aat aac gtc att ccc tgc agg ctt gcc act gtt gct 837
97 Ala Cys Asp Asp Asn Asn Val Ile Pro Cys Arg Leu Ala Thr Val Ala
98 230 235 240
100 tct gga gca gct tcg gga aag ctc ttg caa tac gaa gtt ggt gga cct 885
101 Ser Gly Ala Ala Ser Gly Lys Leu Leu Gln Tyr Glu Val Gly Gly Pro
102 245 250 255
104 aga gtc tgt gtg caa act gca tac ggc gtg gag gtt gag gtg gaa aat 933
105 Arg Val Cys Val Gln Thr Ala Tyr Gly Val Glu Val Glu Val Glu Asn
106 260 265 270 275
108 agt cca tat gat cca gat caa atg gtt ttc atg gat tac aga gat tat 981
109 Ser Pro Tyr Asp Pro Asp Gln Met Val Phe Met Asp Tyr Arg Asp Tyr
110 280 285 290
112 act aac gag aaa gtt cgg agc tta gaa gct gag tat cca acg ttt ctg 1029
113 Thr Asn Glu Lys Val Arg Ser Leu Glu Ala Glu Tyr Pro Thr Phe Leu
114 295 300 305
116 tac gcc atg cct atg aca aag tca aga ctc ttc ttc gag gag aca tgt 1077
117 Tyr Ala Met Pro Met Thr Lys Ser Arg Leu Phe Phe Glu Glu Thr Cys
118 310 315 320
120 ttg gcc tca aaa gat gtc atg ccc ttt gat ttg cta aaa acg aag ctc 1125
121 Leu Ala Ser Lys Asp Val Met Pro Phe Asp Leu Leu Lys Thr Lys Leu
122 325 330 335
124 atg tta aga tta gat aca ctc gga att cga att cta aag act tac gaa 1173
125 Met Leu Arg Leu Asp Thr Leu Gly Ile Arg Ile Leu Lys Thr Tyr Glu
126 340 345 350 355
128 gag gag tgg tcc tat atc cca gtt ggt ggt tcc ttg cca aac acc gaa 1221

```

RAW SEQUENCE LISTING

DATE: 05/08/2003

PATENT APPLICATION: US/09/701,395A

TIME: 14:48:25

Input Set : A:\10817222.app

Output Set: N:\CRF4\05082003\I701395A.raw

```

129 Glu Glu Trp Ser Tyr Ile Pro Val Gly Gly Ser Leu Pro Asn Thr Glu
130          360          365          370
132 caa aag aat ctc gcc ttt ggt gct gcc gct agc atg gta cat ccc gca 1269
133 Gln Lys Asn Leu Ala Phe Gly Ala Ala Ser Met Val His Pro Ala
134          375          380          385
136 aca ggc tat tca gtt gtg aga tct ttg tct gaa gct cca aaa tat gca 1317
137 Thr Gly Tyr Ser Val Val Arg Ser Leu Ser Glu Ala Pro Lys Tyr Ala
138          390          395          400
140 tca gtc atc gca gag ata cta aga gaa gag act acc aaa cag atc aac 1365
141 Ser Val Ile Ala Glu Ile Leu Arg Glu Glu Thr Thr Lys Gln Ile Asn
142          405          410          415
144 agt aat att tca aga caa gct tgg gat act tta tgg cca cca gaa agg 1413
145 Ser Asn Ile Ser Arg Gln Ala Trp Asp Thr Leu Trp Pro Pro Glu Arg
146 420          425          430          435
148 aaa aga cag aga gca ttc ttt ctc ttt ggt ctt gca ctc ata gtt caa 1461
149 Lys Arg Gln Arg Ala Phe Phe Leu Phe Gly Leu Ala Leu Ile Val Gln
150          440          445          450
152 ttc gat acc gaa ggc att aga agc ttc ttc cgt act ttc ttc cgc ctt 1509
153 Phe Asp Thr Glu Gly Ile Arg Ser Phe Phe Arg Thr Phe Phe Arg Leu
154          455          460          465
156 cca aaa tgg atg tgg caa ggg ttt cta gga tca aca tta aca tca gga 1557
157 Pro Lys Trp Met Trp Gln Gly Phe Leu Gly Ser Thr Leu Thr Ser Gly
158          470          475          480
160 gat ctc gtt ctc ttt gct tta tac atg ttc gtc att tca cca aac aat 1605
161 Asp Leu Val Leu Phe Ala Leu Tyr Met Phe Val Ile Ser Pro Asn Asn
162          485          490          495
164 ttg aga aaa ggt ctc atc aat cat ctc atc tct gat cca acc gga gca 1653
165 Leu Arg Lys Gly Leu Ile Asn His Leu Ile Ser Asp Pro Thr Gly Ala
166 500          505          510          515
168 acc atg ata aaa acc tat ctc aaa gta tgattttactt atcaactott 1700
169 Thr Met Ile Lys Thr Tyr Leu Lys Val
170          520
172 aggttttgtgt atatatatgt tgattttatct gaataatcga tcaaagaatg gtatgtgggt 1760
174 tactaggaag ttggaaacaa acatgtatag aatctaagga gtgatcgaaa tggagatgga 1820
176 aacgaaaaga aaaaaatcag tctttgtttt gtggttagtg 1860
179 <210> SEQ ID NO: 2
180 <211> LENGTH: 524
181 <212> TYPE: PRT
182 <213> ORGANISM: Arabidopsis thaliana
184 <400> SEQUENCE: 2
185 Met Glu Cys Val Gly Ala Arg Asn Phe Ala Ala Met Ala Val Ser Thr
186 1 5 10 15
188 Phe Pro Ser Trp Ser Cys Arg Arg Lys Phe Pro Val Val Lys Arg Tyr
189 20 25 30
191 Ser Tyr Arg Asn Ile Arg Phe Gly Leu Cys Ser Val Arg Ala Ser Gly
192 35 40 45
194 Gly Gly Ser Ser Gly Ser Glu Ser Cys Val Ala Val Arg Glu Asp Phe
195 50 55 60
197 Ala Asp Glu Glu Asp Phe Val Lys Ala Gly Gly Ser Glu Ile Leu Phe

```

RAW SEQUENCE LISTING

DATE: 05/08/2003

PATENT APPLICATION: US/09/701,395A

TIME: 14:48:25

Input Set : A:\10817222.app

Output Set: N:\CRF4\05082003\I701395A.raw

```

198 65          70          75          80
200 Val Gln Met Gln Gln Asn Lys Asp Met Asp Glu Gln Ser Lys Leu Val
201          85          90          95
203 Asp Lys Leu Pro Pro Ile Ser Ile Gly Asp Gly Ala Leu Asp His Val
204          100          105          110
206 Val Ile Gly Cys Gly Pro Ala Gly Leu Ala Leu Ala Ala Glu Ser Ala
207          115          120          125
209 Lys Leu Gly Leu Lys Val Gly Leu Ile Gly Pro Asp Leu Pro Phe Thr
210          130          135          140
212 Asn Asn Tyr Gly Val Trp Glu Asp Glu Phe Asn Asp Leu Gly Leu Gln
213 145          150          155          160
215 Lys Cys Ile Glu His Val Trp Arg Glu Thr Ile Val Tyr Leu Asp Asp
216          165          170          175
218 Asp Lys Pro Ile Thr Ile Gly Arg Ala Tyr Gly Arg Val Ser Arg Arg
219          180          185          190
221 Leu Leu His Glu Glu Leu Leu Arg Arg Cys Val Glu Ser Gly Val Ser
222          195          200          205
224 Tyr Leu Ser Ser Lys Val Asp Ser Ile Thr Glu Ala Ser Asp Gly Leu
225          210          215          220
227 Arg Leu Val Ala Cys Asp Asp Asn Asn Val Ile Pro Cys Arg Leu Ala
228 225          230          235          240
230 Thr Val Ala Ser Gly Ala Ala Ser Gly Lys Leu Leu Gln Tyr Glu Val
231          245          250          255
233 Gly Gly Pro Arg Val Cys Val Gln Thr Ala Tyr Gly Val Glu Val Glu
234          260          265          270
236 Val Glu Asn Ser Pro Tyr Asp Pro Asp Gln Met Val Phe Met Asp Tyr
237          275          280          285
239 Arg Asp Tyr Thr Asn Glu Lys Val Arg Ser Leu Glu Ala Glu Tyr Pro
240          290          295          300
242 Thr Phe Leu Tyr Ala Met Pro Met Thr Lys Ser Arg Leu Phe Phe Glu
243 305          310          315          320
245 Glu Thr Cys Leu Ala Ser Lys Asp Val Met Pro Phe Asp Leu Leu Lys
246          325          330          335
248 Thr Lys Leu Met Leu Arg Leu Asp Thr Leu Gly Ile Arg Ile Leu Lys
249          340          345          350
251 Thr Tyr Glu Glu Glu Trp Ser Tyr Ile Pro Val Gly Gly Ser Leu Pro
252          355          360          365
254 Asn Thr Glu Gln Lys Asn Leu Ala Phe Gly Ala Ala Ser Met Val
255          370          375          380
257 His Pro Ala Thr Gly Tyr Ser Val Val Arg Ser Leu Ser Glu Ala Pro
258 385          390          395          400
260 Lys Tyr Ala Ser Val Ile Ala Glu Ile Leu Arg Glu Glu Thr Thr Lys
261          405          410          415
263 Gln Ile Asn Ser Asn Ile Ser Arg Gln Ala Trp Asp Thr Leu Trp Pro
264          420          425          430
266 Pro Glu Arg Lys Arg Gln Arg Ala Phe Phe Leu Phe Gly Leu Ala Leu
267          435          440          445
269 Ile Val Gln Phe Asp Thr Glu Gly Ile Arg Ser Phe Phe Arg Thr Phe
270          450          455          460

```

RAW SEQUENCE LISTING

DATE: 05/08/2003

PATENT APPLICATION: US/09/701,395A

TIME: 14:48:25

Input Set : A:\10817222.app

Output Set: N:\CRF4\05082003\I701395A.raw

```

272 Phe Arg Leu Pro Lys Trp Met Trp Gln Gly Phe Leu Gly Ser Thr Leu
273 465          470          475          480
275 Thr Ser Gly Asp Leu Val Leu Phe Ala Leu Tyr Met Phe Val Ile Ser
276          485          490          495
278 Pro Asn Asn Leu Arg Lys Gly Leu Ile Asn His Leu Ile Ser Asp Pro
279          500          505          510
281 Thr Gly Ala Thr Met Ile Lys Thr Tyr Leu Lys Val
282          515          520
286 <210> SEQ ID NO: 3
287 <211> LENGTH: 956
288 <212> TYPE: DNA
289 <213> ORGANISM: Arabidopsis thaliana
291 <400> SEQUENCE: 3
292 gctctttctc ctctctctct accgatttcc gactccgcct cccgaaatcc ttatccggat 60
293 tctctccgct tcttcgattt aaacgctttt ctgtctgtta cgctcgcgaa gaacggagac 120
294 agaattctcc gattgagaac gatgagagac cggagagcac gagctccaca aacgctatag 180
295 acgctgagta tctggcggtt cgtttgccgg agaaattgga gaggaagaaa tcggagaggt 240
296 ccacttatct aatcgctgct atgttgctga gctttggtat cacttctatg gctgttatgg 300
297 ctgtttacta cagattctct tggcaaattg agggaggtga gatctcaatg ttggaaatgt 360
298 ttggtacatt tgctctctct gttggtgctg ctgttggtat ggaattctgg gcaagatggg 420
299 ctcatagagc tctgtggcac gcttctctat ggaatatgca tgagtcacat cacaaccaa 480
300 gagaaggacc gtttgagcta aacgatgttt ttgctatagt gaacgctggt ccagcgattg 540
301 gtctctctc ttatggattc ttcaataaag gactcgttcc tggctctctgc tttggcgccg 600
302 ggtaggcata aacggtgttt ggaatgcct acatgtttgt ccacgatggt ctcgtcaca 660
303 agcgtttccc tgtaggctcc atcgccgacg tcccttacct ccgaaaggtc gccgcgctc 720
304 accagctaca tcacacagac aagttcaatg gtgtaccata tggactgttt cttggaccca 780
305 aggaattgga agaagttgga ggaaatgaag agttagataa ggagattagt cggagaatca 840
306 aatcatacaa aaaggcctcg ggctccgggt cgagttcgag ttcttgactt taaacaagtt 900
307 ttaaatecca aattcttttt ttgtctctg tcattatgat catcttaaga cggctct 956
310 <210> SEQ ID NO: 4
311 <211> LENGTH: 294
312 <212> TYPE: PRT
313 <213> ORGANISM: Arabidopsis thaliana
315 <400> SEQUENCE: 4
316 Ser Phe Ser Ser Ser Ser Thr Asp Phe Arg Leu Arg Leu Pro Lys Ser
317 1          5          10          15
319 Leu Ser Gly Phe Ser Pro Ser Leu Arg Phe Lys Arg Phe Ser Val Cys
320          20          25          30
322 Tyr Val Val Glu Glu Arg Arg Gln Asn Ser Pro Ile Glu Asn Asp Glu
323          35          40          45
325 Arg Pro Glu Ser Thr Ser Ser Thr Asn Ala Ile Asp Ala Glu Tyr Leu
326          50          55          60
328 Ala Leu Arg Leu Ala Glu Lys Leu Glu Arg Lys Lys Ser Glu Arg Ser
329 65          70          75          80
331 Thr Tyr Leu Ile Ala Ala Met Leu Ser Ser Phe Gly Ile Thr Ser Met
332          85          90          95
334 Ala Val Met Ala Val Tyr Tyr Arg Phe Ser Trp Gln Met Glu Gly Gly
335          100          105          110
337 Glu Ile Ser Met Leu Glu Met Phe Gly Thr Phe Ala Leu Ser Val Gly

```

RAW SEQUENCE LISTING ERROR SUMMARY
PATENT APPLICATION: US/09/701,395A

DATE: 05/08/2003
TIME: 14:48:26

Input Set : A:\10817222.app
Output Set: N:\CRF4\05082003\I701395A.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:13; N Pos. 366,367,368,369,370,371,372,373,374,375,376,377,378,379,380
Seq#:13; N Pos. 381,382,383,384,385,386,387,388,389,390,391,392,393,394,395
Seq#:13; N Pos. 396,397,398,399,400,401,402,403,404,405,406,407,408,409,410
Seq#:13; N Pos. 411,412,413,414,415,416,417,418,419,420,421,422,423,424,425
Seq#:13; N Pos. 426,427,428,429,430,431,432,433,434,435,436,437,438,439,440
Seq#:13; N Pos. 441,442,443,444,445,446,447,448,449,450,451,452,453,454,455
Seq#:13; N Pos. 456,457,458,459,460,461,462,463,464,465,466,467,468,469,470
Seq#:13; N Pos. 471,472,473,474,475,476,477,478,479,480,481,482,483,484,485
Seq#:13; N Pos. 486,487,488,489,490,491,492,493,494,495,496,497,498,499,500
Seq#:13; N Pos. 501,502,503,504,505,506,507,508,509,510,511,512,513,514,515
Seq#:13; N Pos. 516,517,518,519,520,521,522,523,524,525,526,527,528,529,530
Seq#:13; N Pos. 531,532,533,534,535,536,537,538,539,540,541,542,543,544,545
Seq#:13; N Pos. 546,547,548,549,550,551,552,553,554,555,556,557,558,559,560
Seq#:13; N Pos. 561,562,563,564,565,566,567,568,569,570,571,572,573,574,575
Seq#:13; N Pos. 576,577,578,579,580,581,582,583,584,585,586,587,588,589,590
Seq#:13; N Pos. 591,592,593,594,595,596,597,598,599,600,601,602,603,604,605
Seq#:13; N Pos. 606,607,608,609,610,611,612,613,614,615,616,617,618,619,620
Seq#:13; N Pos. 621,622,623,624,625,626,627,628,629,630,631,632,633,634,635
Seq#:13; N Pos. 636,637,638,639,640,641,642,643,644,645,646,647,648,649,650
Seq#:13; N Pos. 651,652,653,654,655,656,657,658,659,660,661,662,663,664,665
Seq#:13; N Pos. 666,667,668,669,670,671,672,673,674,675,676,677,678,679,680
Seq#:20; Xaa Pos. 6,10,12,14,15,16,17,18,22,23,24,25,27,28,29,30,31,32,33
Seq#:20; Xaa Pos. 34,35,36,38,39,41,42,43,44,45,46,47,48,49,50,51,53,54,55
Seq#:20; Xaa Pos. 57,61,77,80,83,84,90,112,116,117,144,149,150,153,155,156
Seq#:20; Xaa Pos. 157,158,162,178,183,190,191,194,195,197,201,202,204,205
Seq#:20; Xaa Pos. 206,208,211,212,216,227,236,250,251,256,257,261,267,270
Seq#:20; Xaa Pos. 273,274,280,281,282,295,310,311,312,319,323,328,331,337
Seq#:20; Xaa Pos. 339,345,349,356,362,380,386,387,390,392,393,394,395,396
Seq#:20; Xaa Pos. 398,399,402,403,406,433,446,448,459,464,465,473,480,488
Seq#:20; Xaa Pos. 491,492,498,501,502
Seq#:21; Xaa Pos. 321
Seq#:24; N Pos. 7,8,9,10,11,1020,1180,1181,1330
Seq#:25; Xaa Pos. 336
Seq#:26; Xaa Pos. 491
Seq#:34; N Pos. 565,569
Seq#:48; Xaa Pos. 336

VERIFICATION SUMMARY

DATE: 05/08/2003

PATENT APPLICATION: US/09/701,395A

TIME: 14:48:26

Input Set : A:\10817222.app

Output Set: N:\CRF4\05082003\I701395A.raw

L:655 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:13 after pos.:360
M:341 Repeated in SeqNo=13
L:1641 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:20 after pos.:0
M:341 Repeated in SeqNo=20
L:1809 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:21 after pos.:320
L:2024 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:24 after pos.:0
M:341 Repeated in SeqNo=24
L:2120 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:25 after pos.:320
L:2238 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:26 after pos.:480
L:2527 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:34 after pos.:540
L:3366 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:48 after pos.:320